Attorney Docket No. YOR920030088US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Y. Diao et al.
Docket No.: YOR920030088US1
Serial No.: 10/648,179
Filing Date: August 26, 2003
Group: 2123
Examiner: Juan Carlos Ochoa

Title: Methods and Systems for Model-Based Management Using Abstract Models

---

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Appellants (hereinafter referred to as "Appellants") hereby appeal the final rejection of claims 1-33 of the above-identified application.

REAL PARTY IN INTEREST

The present application is assigned to International Business Machines Corporation, as evidenced by an assignment recorded August 26, 2003 in the U.S. Patent and Trademark Office at Reel 14442, Frame 276. The assignee, International Business Machines Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no known related appeals or interferences.

STATUS OF CLAIMS

Claims 1-33 stand finally rejected under 35 U.S.C. §102(b).

STATUS OF AMENDMENTS

There has been no amendment filed subsequent to the final rejection.

SUMMARY OF CLAIMED SUBJECT MATTER

Independent claim 1 recites a method of constructing a model representative of a resource for use in managing a service associated with the resource, comprising the steps of associating a resource abstract model with the resource, wherein the resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

An illustrative embodiment of the present invention includes an exemplary method of constructing a model (e.g., element 180 in FIG. 1) representative of a resource (e.g., 115, 120, 125, 140, 145, 150, 155 in FIG. 1) for use in managing a service associated with the resource, described in the specification at, for example, page 9, lines 17-25, and page 10, lines 8-21, both with reference to FIG. 1. The exemplary method includes a step of associating a resource abstract model (e.g., 110 and 135 in FIG. 1) with the resource, described in the specification at, for example, page 12, line 15 to page 13, line 2, with reference to FIG. 4. As described in the specification at, for example, page 8, line 8, to page 9, line 9, the resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered. The exemplary method also includes a step of constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model, described in the specification at, for example, page 13, lines 12-27, with reference to FIG. 5.

Independent claim 11 recites an apparatus for constructing a model representative of a resource for use in managing a service associated with the resource. The apparatus comprises a memory and at least one processor coupled to the memory. The processor is operative to automatically determine, via a resource abstract model, a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered. The processor is further operative to construct the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

An illustrative embodiment of the present invention includes an exemplary apparatus for constructing a model (e.g., 180 in FIG. 1) representative of a resource (e.g., 115, 120, 125, 140, 145, 150, 155 in FIG. 1) for use in managing a service associated with the resource, described in the specification at, for example, page 9, lines 17-25, and page 10, lines 8-21, both with reference to FIG. 1. As described in the specification at, for example, page 15, line 23 to page 16, line 15, with reference to FIG. 8, the apparatus comprises a memory (e.g., 804) and at least one processor (e.g., 802) coupled to the memory. The processor is operative to automatically determine, via a resource abstract model (e.g., 110 and 135 in FIG. 1), a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and to construct the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model. See the specification at, for example, for example, page 8, line 8, to page 9, line 9; page 12, line 15 to page 13, line 2, with reference to FIG. 4; and page 13, lines 12-27, with reference to FIG. 5.

Independent claim 20 recites an article of manufacture for constructing a model representative of a resource for use in managing a service associated with the resource. The article of manufacture comprises a machine readable medium containing one or more programs. When executed, the one or more programs implement the steps of automatically determining, via a resource abstract model, a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance

with the resource abstract model.

An illustrative embodiment of the present invention includes an exemplary article of manufacture for constructing a model (e.g., element 180 in FIG. 1) representative of a resource (e.g., 115, 120, 125, 140, 145, 150, 155 in FIG. 1) for use in managing a service associated with the resource, described in the specification at, for example, page 9, lines 17-25, and page 10, lines 8-21, both with reference to FIG. 1. As described in the present specification at, for example, page 17, lines 6-10, the article of manufacture comprises a machine readable medium containing one or more programs. When executed, the one or more programs implement the steps of automatically determining, via a resource abstract model (e.g., 110 and 135 in FIG. 1), a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model. See the specification at, for example, for example, page 8, line 8, to page 9, line 9; page 12, line 15 to page 13, line 2, with reference to FIG. 4; and page 13, lines 12-27, with reference to FIG. 5.

Independent claim 27 recites a method of providing resource management services, including a step of deploying one or more resource abstract models in association with one or more resources. Each of the one or more resource abstract models is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered. The exemplary method also includes steps of, based on the one or more reduced sets of resource metrics obtained in accordance with the one or more resource abstract models, constructing one or more models representative of the one or more resources; and using the one or more constructed models to manage the one or more resources.

An exemplary method of providing resource management services includes a step of deploying one or more resource abstract models in association with one or more resources, as described in the specification at, for example, page 9, lines 17-25, and page 10, lines 8-21, both with reference to FIG. 1; page 12, line 15 to page 13, line 2, with reference to FIG. 4; and page 14, lines 16-21, with reference to FIG. 7. Each of the one or more resource abstract models is configured to automatically determine a set of resource metrics to be used to construct a model (e.g., 180 in FIG. 1)

representative of the resource such that a reduced set of resource metrics is considered. See the specification at, for example, for example, page 8, line 8, to page 9, line 9. The exemplary method also includes steps of, based on the one or more reduced sets of resource metrics obtained in accordance with the one or more resource abstract models, constructing one or more models representative of the one or more resources; and using the one or more constructed models to manage the one or more resources. See the specification at, for example, page 13, lines 12-27, with reference to FIG. 5.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1- 33 are rejected under 35 U.S.C. §102(b) as being unpatentable over Bigus et al., "AutoTune: A Generic Agent for Automated Performance Tuning," *Practical Application of Intelligent Agents and Multi Agent Technology*, 2000 (hereinafter Bigus).

ARGUMENT

Claims 1-5, 7-14, 16-23, 25-29 and 31-33

With regard to the §102 rejection, Appellants initially note that a claim is anticipated "only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). Moreover, the cited reference must show the "identical invention . . . in as complete detail as is contained in the . . . claim," citing *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 1236, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989). See also MPEP §2131.

Independent claim 1 is directed to a method of constructing a model representative of a resource for use in managing a service associated with the resource, comprising the steps of associating a resource abstract model with the resource; wherein the resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model. It should be noted that the present

specification at page 9, lines 7-8, indicates that a resource abstract model "may be considered a computer readable description of one or more metrics."

The Examiner contends that the steps recited above are taught by Bigus at page 5, last paragraph, and page 11, lines 6-10. Page 5, last paragraph, of Bigus recites, in relevant part:

> Our starting point is the target's system model. As shown in Figure 3, the system model is an abstraction of the target that outputs service levels given inputs for workload, configuration, and settings of tuning controls. This model can be constructed using various learning approaches that enable different control algorithms to be employed. In our current prototype, the system model is obtained by training a neural network based on measured values of the controlled target over a wide range of workloads and tuning controls.

Page 11, lines 1-10, of Bigus, which includes the section cited by the Examiner, recite:

> The core NeuralPredictionAgent . . . . uses an AbleImport to read training and test data from text files, uses two AbleFilters, one to pre-process the data and one to post-process the data, and a back propagation neural network to perform the regression function. . . . The user need only specify the source data file (and a corresponding meta-data file). The neural prediction agent then scans the source data and automatically generates the scaling and transformation templates used by the AbleFilters to pre- and post- process the data going into and out of the neural network. Based on the number of inputs and output fields and their data representation, the neural network architecture is automatically configured.

The above-quoted portions of Bigus appear to be directed to constructing a system model, which is an abstraction of the target that outputs service levels given certain inputs, by training a neural network based on measured values of the controlled target. A neural prediction agent scans a user-specified source data file and automatically generates the scaling and transformation templates used to pre- and post- process the data going into and out of the neural network. Based on the number of inputs and output fields and their data representation, a neural network architecture is automatically configured.

Appellants thus respectfully submit that the above-quoted portions of Bigus fail to teach or suggest at least the limitation recited in claim 1 wherein a resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of

the resource such that a reduced set of resource metrics is considered. Accordingly, Appellants assert that claim 1 is patentable over Bigus.

In the final Office Action at page 8, third and fourth paragraphs, the Examiner appears mischaracterize the above argument. Appellants are not arguing that Bigus fails to teach a limitation allegedly recited in claim 1 wherein a resource abstract model "may be considered a computer-readable description of one or more metrics." Moreover, Appellants disagree with the Examiner's contention that the above argument is an impermissible attempt to read limitations from the specification into the claims.

Rather, Appellants respectfully submit that Bigus fails to teach or suggest the limitations recited in claim 1 wherein a method of constructing a model representative of a resource for use in managing a service associated with the resource comprises the steps of associating a resource abstract model with the resource; wherein the resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

Appellants note that, where a definition is provided by the applicant for a term, either explicitly or by implication (i.e., according to the usage of the term in the context in the specification), that definition will control interpretation of the term as it is used in the claim. See *Vitronics Corp. v. Conceptronic Inc.*, 90 F.3d 1576, 1583, 39 USPQ2d 1573, 1577 (Fed. Cir. 1996); see generally *Phillips v. AWH Corp.*, 415 F.3d 1303, 75 USPQ2d 1321 (Fed. Cir. 2005) (*en banc*).

Accordingly, Appellants are not attempting to read limitations from the specification into the claims, but rather are interpreting the claim in light of the specification. See, e.g., *Constant v. Advanced Micro-Devices, Inc.*, 848 F.2d 1560, 1572, 7 USPQ2d 1057, 1065 (Fed. Cir. 1988) (If "words that are used in the claims [are] defined in the specification," these definitions from the specification "must be imported into the claims to give meaning to disputed terms.")

In the present Office Action at page 8, last paragraph, the Examiner further argues that Bigus at page 9, section 3.2, line 1 ("In the ABLE framework, an agent is an autonomous software

component") discloses the limitation allegedly argued by Appellants. Appellants respectfully submit this disclosure that an agent is autonomous software component fails to teach, or even suggest, the resource abstract model recited in claim 1, which, as noted above, is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered.

Independent claims 11, 20 and 27 include limitations similar to those of claim 1, and are therefore believed allowable for reasons similar to those described above with reference to claim 1.

Dependent claims 2-5, 7-10, 12-14, 16-19, 21-23, 25, 26, 28, 29 and 31-33 are believed patentable at least by virtue of their respective dependence on independent claims 1, 11, 20 and 27.

Claims 6, 15, 24 and 30

In addition to being patentable by virtue of their respective dependence on independent claims 1, 11, 20 and 27, dependent claims 6, 15, 24 and 30 are believed to define separately patentable subject matter. For example, dependent claim 6 includes a limitation directed to obtaining a topology of one or more resources used to deliver one or more services associated with the one or more service level agreements, including the resource for which the model is being constructed, for use in constructing the model representative of the resource.

In an illustrative embodiment described in the specification at page 7, lines 8-11, a configuration database specifies the topology of resources used to deliver the services described in the service level agreement. A topology refers to the minimal set of resources that may be used in service delivery. More detailed topologies specify the resources actually used and the flows between them. For example, as described in the specification at page 15, lines 1-3, a service topology for an exemplary service level agreement may be a single computer running DB2.

In formulating the rejection of claim 6 at page 3, fourth paragraph, of the final Office Action, the Examiner argues that the aforementioned limitation is met by Bigus at page 18, last paragraph. Specifically, the Examiner states that "per the topology definition in (application description page 7, 4th paragraph), Examiner interprets 'scheduling different classes of customers on a set of distributed

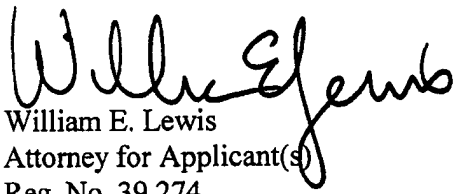
heterogeneous servers to globally minimize a linear function of the per-class mean response times' as minimal set of resources that may be used in service delivery and the flows between them."

Appellants respectfully disagree with the Examiner's interpretation. The relied-upon portion of Bigus does not disclose any minimal set of resources that may be used in service delivery and the flows between them; rather, it is directed to optimizing scheduling of customers on a predetermined set of server resources so as to minimize a response time metric. In other words, there is no teaching or suggestion that the "set of distributed heterogeneous servers" is a minimal set of resources that may be used in service delivery.

Accordingly, Appellants believe that the relied-upon portion of Bigus fails to disclose at least the limitation of dependent claim 6 directed to obtaining a topology of one or more resources used to deliver one or more services associated with the one or more service level agreements. Dependent claims 15, 24 and 30 include limitations similar to those recited in claim 6, and are hence believed patentable for reasons similar to those identified above.

In view of the above, Appellants believe that claims 1-33 are in condition for allowance, and respectfully request reversal of the §102 rejection.

Respectfully submitted,



Date: October 15, 2008

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

APPENDIX

1. A method of constructing a model representative of a resource for use in managing a service associated with the resource, comprising the steps of:

associating a resource abstract model with the resource, wherein the resource abstract model is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and

constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

2. The method of claim 1, wherein the constructed model comprises a quantitative model.

3. The method of claim 1, wherein the resource abstract model is constructed by at least one individual with expertise associated with the resource.

4. The method of claim 1, further comprising the step of obtaining one or more service level metrics for use in constructing the model representative of the resource.

5. The method of claim 4, wherein the one or more service level metrics are obtainable from one or more service level agreements.

6. The method of claim 5, further comprising the step of obtaining a topology of one or more resources used to deliver one or more services associated with the one or more service level agreements, including the resource for which the model is being constructed, for use in constructing the model representative of the resource.

7. The method of claim 1, wherein the resource is an element of an autonomic computing environment.

8. The method of claim 1, wherein the constructed model is useable for at least one of: (i) reporting one or more service level metrics; (ii) automating service level compliance; (iii) permitting a service provider to manage one or more service on demand; and (iv) generating one or more notifications related to automated service level enforcement.

9. The method of claim 1, further comprising the step of checking the accuracy of the constructed model.

10. The method of claim 9, wherein the accuracy checking step comprises use of change point detection.

11. Apparatus for constructing a model representative of a resource for use in managing a service associated with the resource, comprising:

a memory; and

at least one processor coupled to the memory and operative to: (i) automatically determining, via a resource abstract model, a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and (ii) constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

12. The apparatus of claim 11, wherein the constructed model comprises a quantitative model.

13. The apparatus of claim 11, wherein the at least one processor is further operative to obtain one or more service level metrics for use in constructing the model representative of the resource.

14. The apparatus of claim 13, wherein the one or more service level metrics are obtainable from one or more service level agreements.

15. The apparatus of claim 14, wherein the at least one processor is further operative to obtain a topology of one or more resources used to deliver one or more services associated with the one or more service level agreements, including the resource for which the model is being constructed, for use in constructing the model representative of the resource.

16. The apparatus of claim 11, wherein the resource is an element of an autonomic computing environment.

17. The apparatus of claim 11, wherein the constructed model is useable for at least one of: (i) reporting one or more service level metrics; (ii) automating service level compliance; (iii) permitting a service provider to manage one or more service on demand; and (iv) generating one or more notifications related to automated service level enforcement.

18. The apparatus of claim 11, wherein the at least one processor is further operative to check the accuracy of the constructed model.

19. The apparatus of claim 18, wherein the accuracy checking operation comprises use of change point detection.

20. An article of manufacture for constructing a model representative of a resource for use in managing a service associated with the resource, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

automatically determining, via a resource abstract model, a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and

constructing the model representative of the resource based on the reduced set of resource metrics obtained in accordance with the resource abstract model.

21. The article of claim 20, wherein the constructed model comprises a quantitative model.

22. The article of claim 20, further comprising the step of obtaining one or more service level metrics for use in constructing the model representative of the resource.

23. The article of claim 22, wherein the one or more service level metrics are obtainable from one or more service level agreements.

24. The article of claim 20, further comprising the step of obtaining a topology of one or more resources used to deliver one or more services associated with the one or more service level agreements, including the resource for which the model is being constructed, for use in constructing the model representative of the resource.

25. The article of claim 20, wherein the constructed model is useable for at least one of: (i) reporting one or more service level metrics; (ii) automating service level compliance; (iii) permitting a service provider to manage one or more service on demand; and (iv) generating one or more notifications related to automated service level enforcement.

26. The article of claim 20, further comprising the step of checking the accuracy of the constructed model.

27. A method of providing resource management services, comprising the steps of:

deploying one or more resource abstract models in association with one or more resources, wherein each of the one or more resource abstract models is configured to automatically determine a set of resource metrics to be used to construct a model representative of the resource such that a reduced set of resource metrics is considered; and

based on the one or more reduced sets of resource metrics obtained in accordance with the one or more resource abstract models, constructing one or more models representative of the one or more resources; and

using the one or more constructed models to manage the one or more resources.

28. The method of claim 27, further comprising the step of obtaining one or more service level metrics for use in constructing the one or more models representative of the one or more resources.

29. The method of claim 28, wherein the one or more service level metrics are obtainable from one or more service level agreements.

30. The method of claim 27, further comprising the step of obtaining a topology of the one or more resources used to deliver one or more services associated with the one or more service level agreements, for use in constructing the one or more models representative of the resource.

31. The method of claim 27, wherein the resource is an element of an autonomic computing environment.

32. The method of claim 27, wherein the one or more constructed models are useable for at least one of: (i) reporting one or more service level metrics; (ii) automating service level compliance; (iii) permitting a service provider to manage one or more service on demand; and (iv) generating one or more notifications related to automated service level enforcement.

33. The method of claim 27, further comprising the step of checking the accuracy of the one or more constructed models.

EVIDENCE APPENDIX

Attached hereto is a Declaration of Prior Invention Under 37 C.F.R. §1.131 and accompanying Exhibit 1, originally submitted on October 17, 2007.

Docket No. YOR920030088US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Y. Diao et al.
Docket No.: YOR920030088US1
Serial No.: 110/648,179
Filing Date: August 26, 2003
Group: 2123
Examiner: Juan Carlos Ochoa

Title: Methods and Systems for Model-Based Management Using Abstract Models

---

DECLARATION OF PRIOR INVENTION UNDER 37 C.F.R. §1.131

We, the undersigned, hereby declare and state as follows:

1. We are named joint inventors of the invention that is the subject of the above-referenced U.S. patent application. We have assigned our respective interests in the patent application to International Business Machines Corporation ("IBM").

2. The invention falling within the scope of the claims in the present application was conceived and reduced to practice at some time prior to July 3, 2002.

3. On or about July 3, 2002, an IBM proprietary presentation named "YixinDemoShots070302.PRZ" describing the invention was prepared. A copy of the relevant portion of this presentation, consisting of slides 6-16 thereof, is attached hereto as Exhibit 1.

4. The slides attached hereto as Exhibit 1 demonstrate an actual reduction to practice of the invention in the form of experimental results generated using an exemplary embodiment according to the inventive principles.

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: Oct. 12, 2007

Yixin Diao

Date: _____

Denise Y. Dyko

Date: _____

Frank N. Eskensen

Date: _____

Joseph L. Hellerstein

Date: _____

Alexander Keller

Date: _____

Lisa F. Spainhower

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: _____

Yixin Diao

Date: 10/11/2007

Denise Y. Dyko

Date: _____

Frank N. Eskensen

Date: _____

Joseph L. Hellerstein

Date: _____

Alexander Keller

Date: _____

Lisa F. Spainhower

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: _____ Yixin Diao

Date: _____ Denise Y. Dyko

Date: 10/12/07 Frank N. Eskesen

Date: _____ Joseph L. Hellerstein

Date: _____ Alexander Keller

Date: _____ Lisa F. Spainhower

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: _____ Yixin Diao

Date: _____ Denise Y. Dyko

Date: _____ Frank N. Eskensen

Date: 10/11/2007 Joseph L. Hellerstein

Date: _____ Alexander Keller

Date: _____ Lisa F. Spainhower

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: _____

Yixin Diao

Date: _____

Denise Y. Dyko

Date: _____

Frank N. Eskensen

Date: _____

Joseph L. Hellerstein

Date: 10/11/2007

Alexander Keller

Date: _____

Lisa F. Spainhower

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: _____ Yixin Diao

Date: _____ Denise Y. Dyko

Date: _____ Frank N. Eskensen

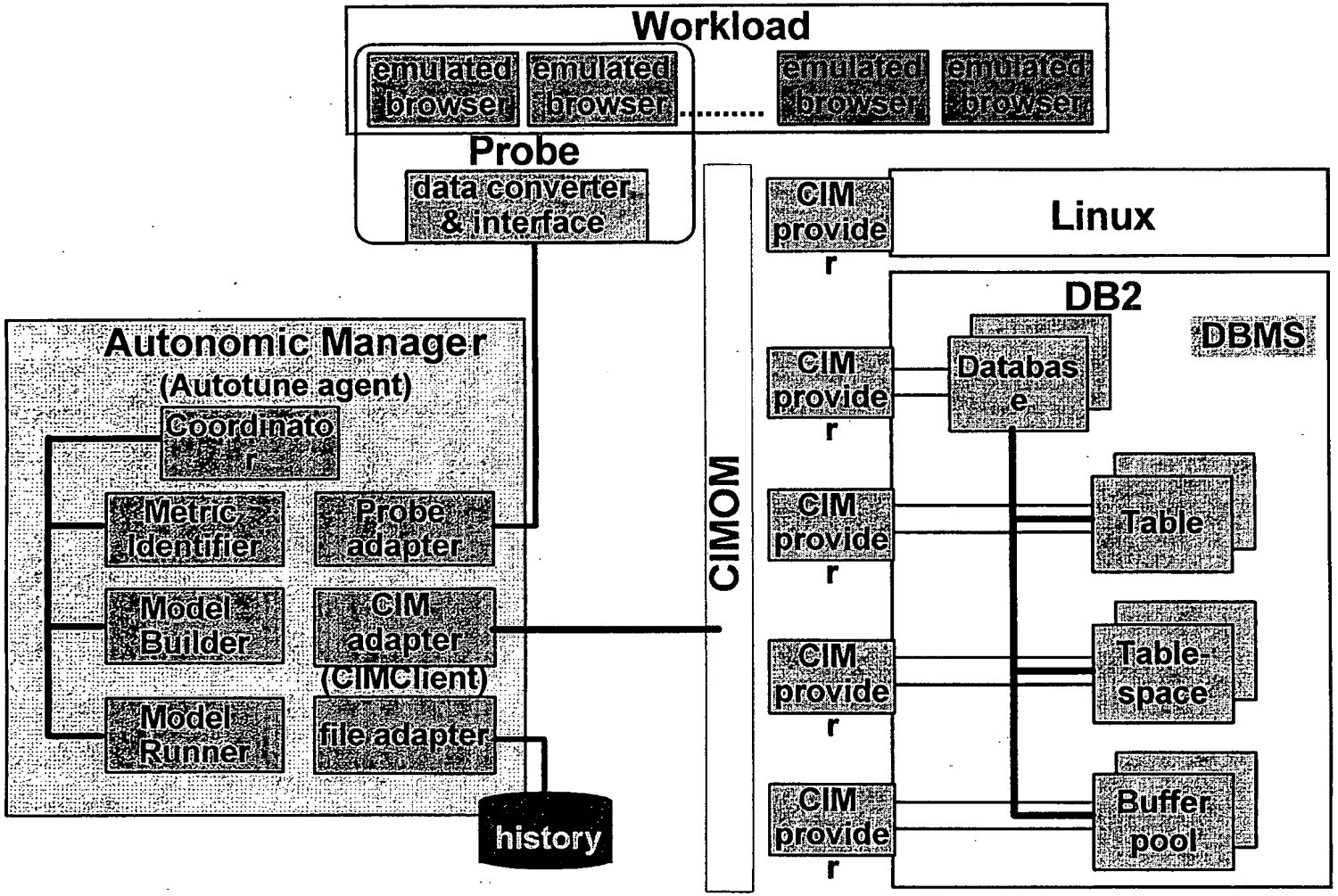
Date: _____ Joseph L. Hellerstein

Date: _____ Alexander Keller

Date: 10-11-07 Lisa F. Spainhower

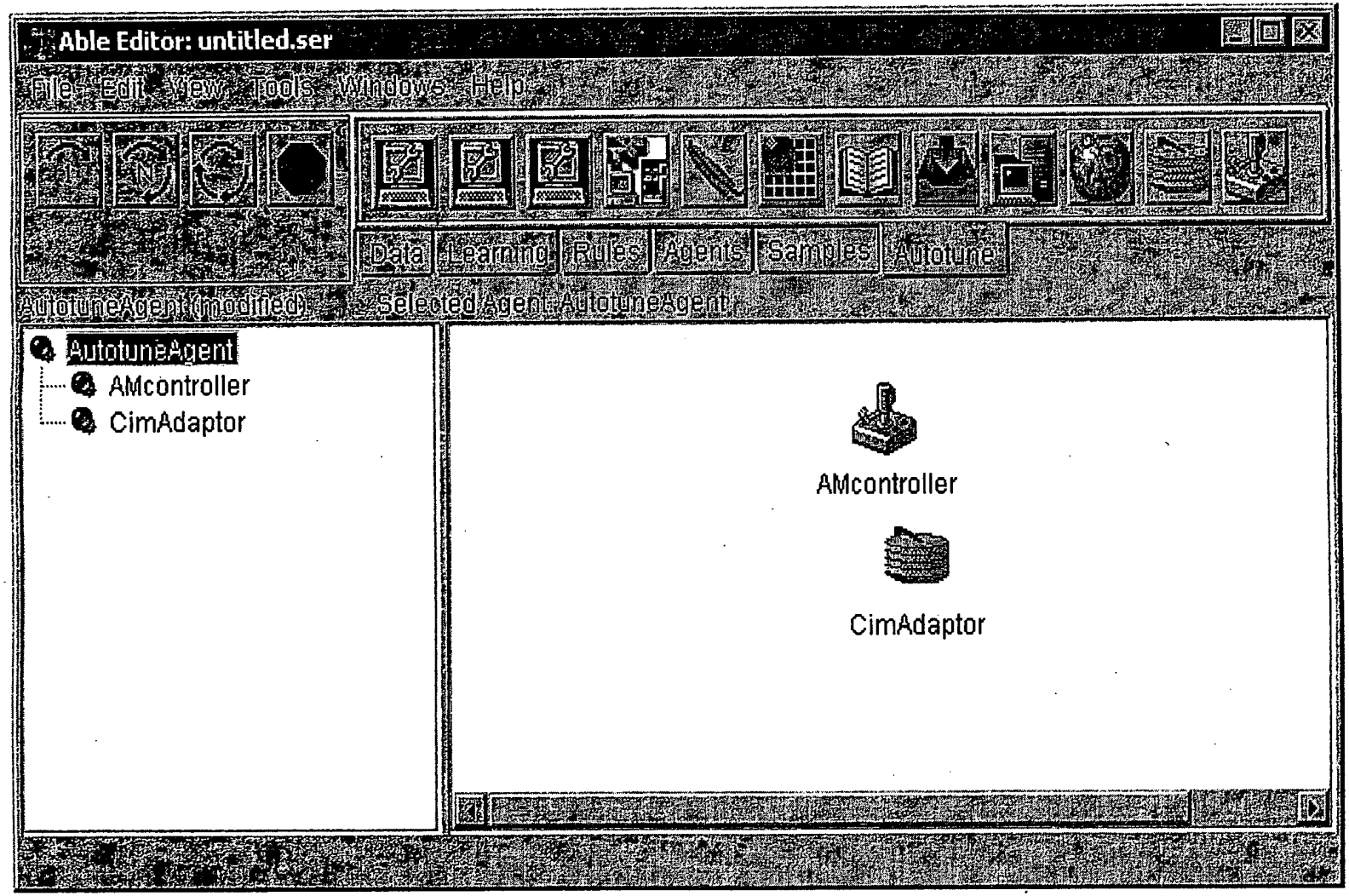
EXHIBIT 1

Architecture for demo: DB2 on Linux platform with TPC-W workload

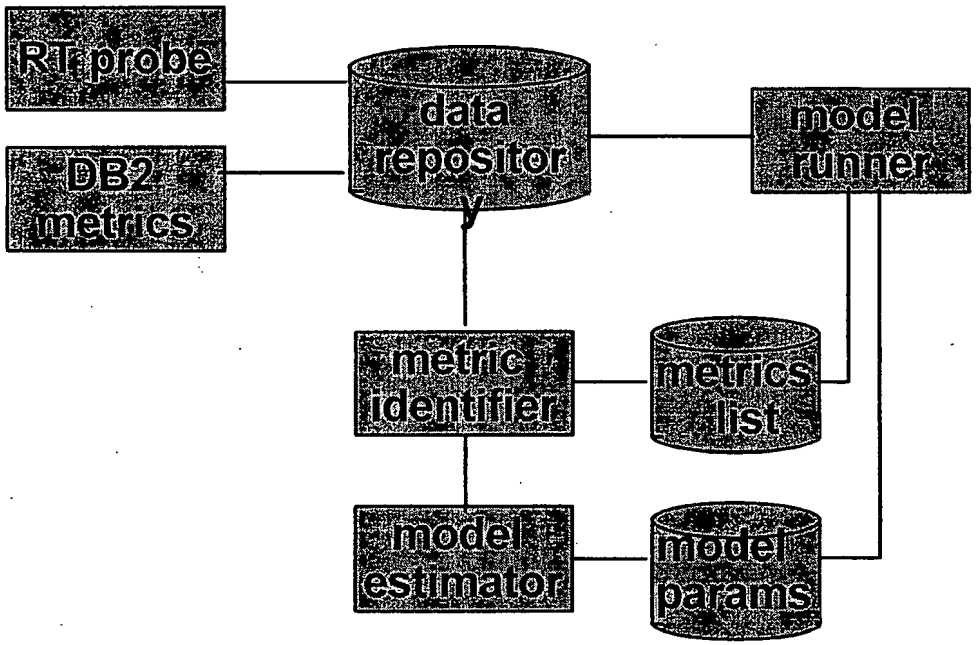


show what we build in one colour -e.g. CIM provider, all stuff in AM, and probe ir

Building AM with AutoTune

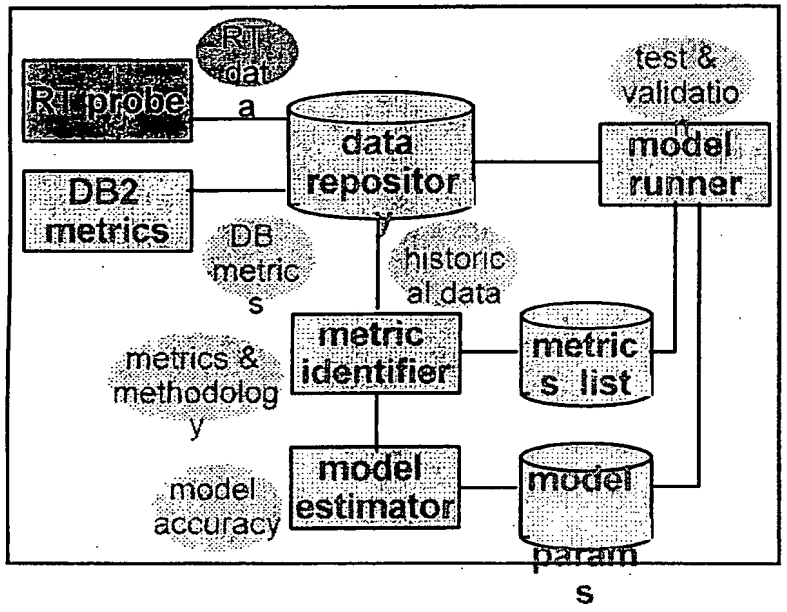


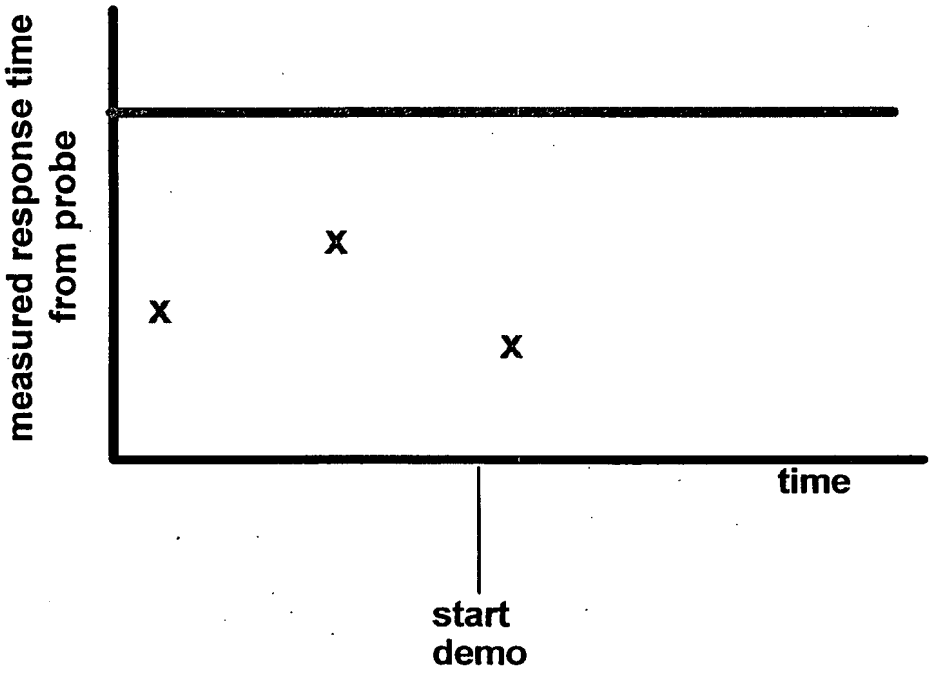
Demo Flow


RTprobe
DB2 metrics
data repository
model runner
metric identifier
metrics list
model estimator
model params

Metric Discovery - demo flow 1

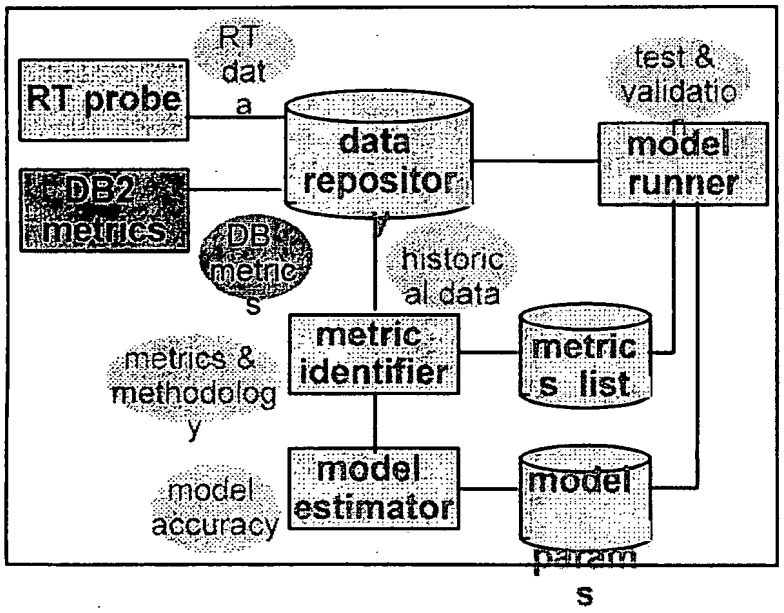
- Show response time as is being collected from probe
  - e.g. one point every 20 sec (20 min for model build) from RT probes
    - (avg over several TPC-W emulated browsers)
  - need to be running/collecting data before demo starts or read in from file
    - require connection to machine w DB server and EBs?

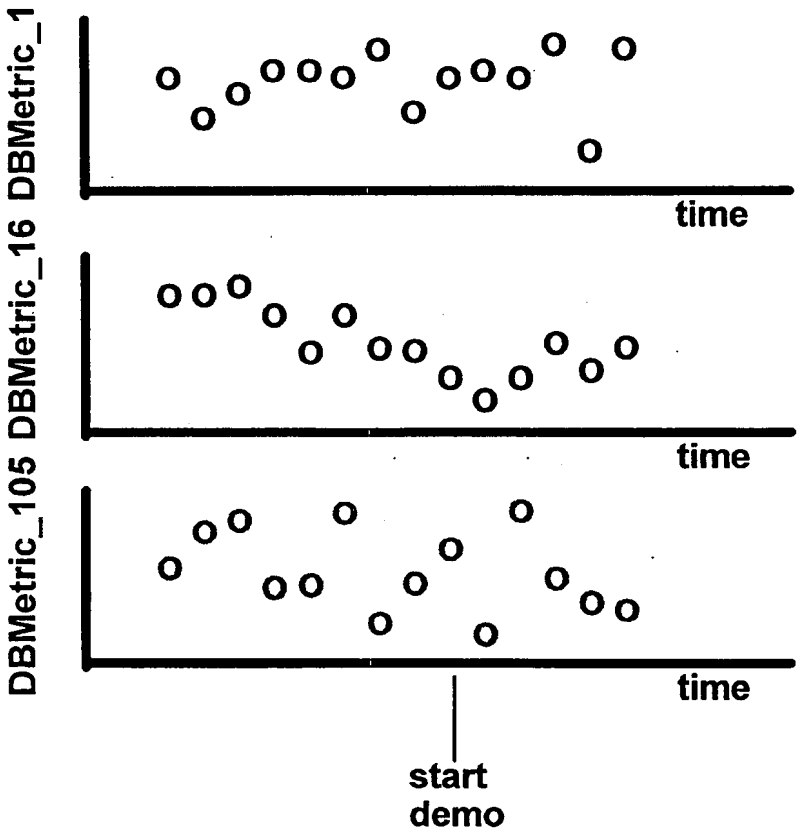




Internal metrics from DB2 from CIM providers - demo flow 2

- Show internal DB2 metrics being collected by Autonomic Manager at rapid intervals
  - ➢ Demonstrates the operational capability of the abstract model
    - –interface between AM Autotune agent and DB2 CIM provider
  - ➢ Show time series of several (2-4) metrics
    - –every 2 sec (2 min for model build)

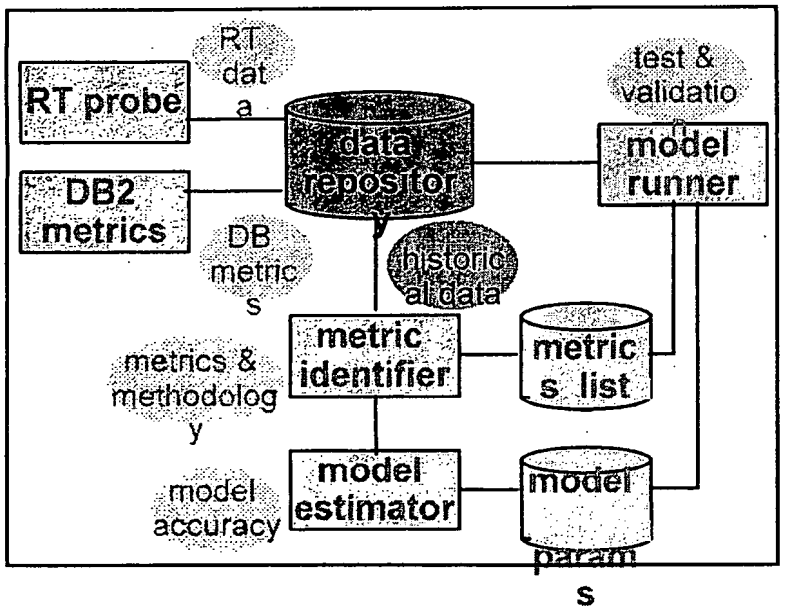


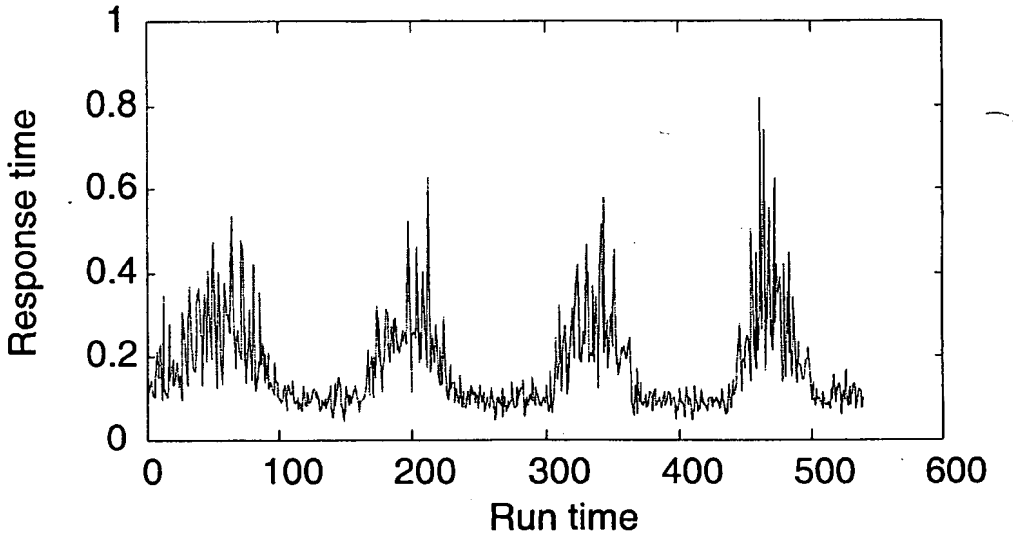


Getting Properties from CIMOM

Metric discovery: done with historical data - demo flow 3

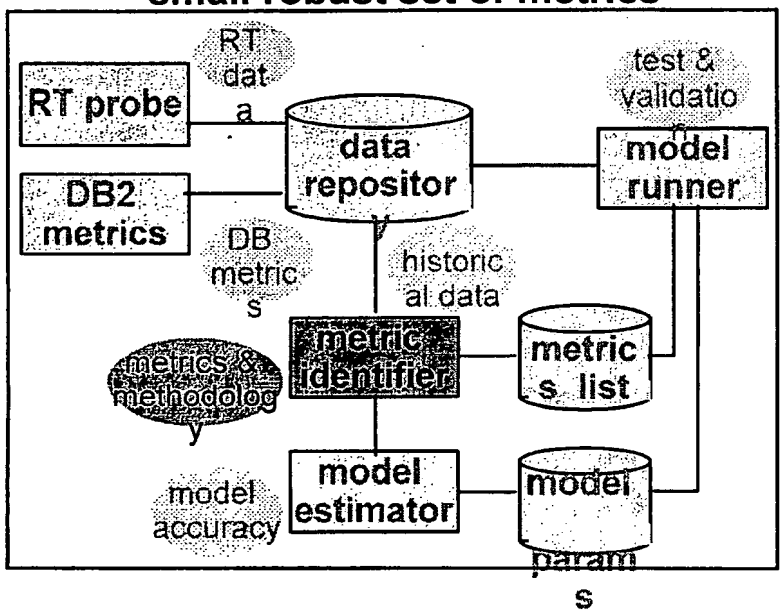
- Model build using historical data
  - Data collected over 10-20 hours from response time probes and DB
    - –mixed workload (vary buy/browse) and varied (# emulated browsers)
    - –training time length a concern?
  - RT every 20 min (RT averaged over k min, k ~ 2-5 min) and DB metrics every k min
    - –cannot use the shorter intervals in real time part of demo as averages over 2 sec will have a lot of noise and RT probe sampling every 20 sec is unrealistic



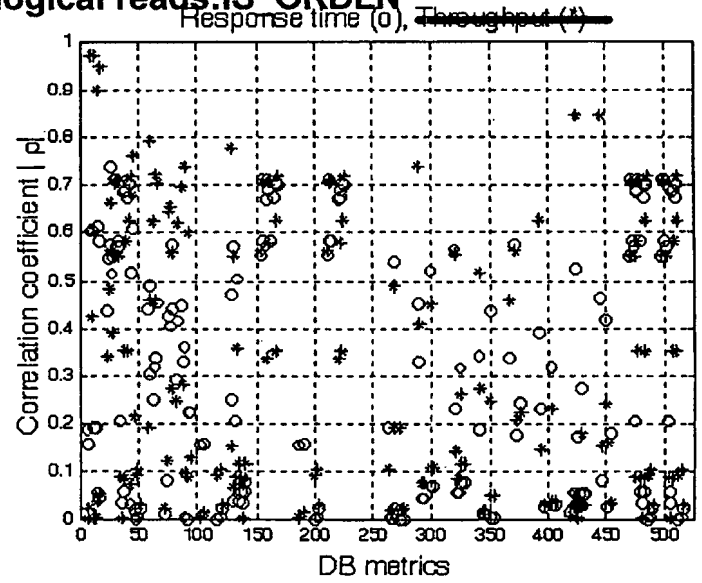


Metric discovery: identify relevant DB metrics- demo flow 4

- Metric identification
  - DB metrics are a mix of cumulative stats and actual snap shots
    - –use metrics as is and 1st order differences in search for correlations to response time
  - Report relevant metrics: high (and ~ independent) correlations to RT
    - –small robust set of metrics

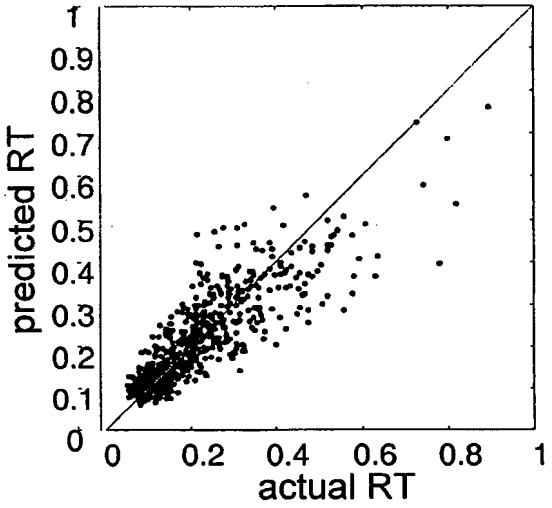


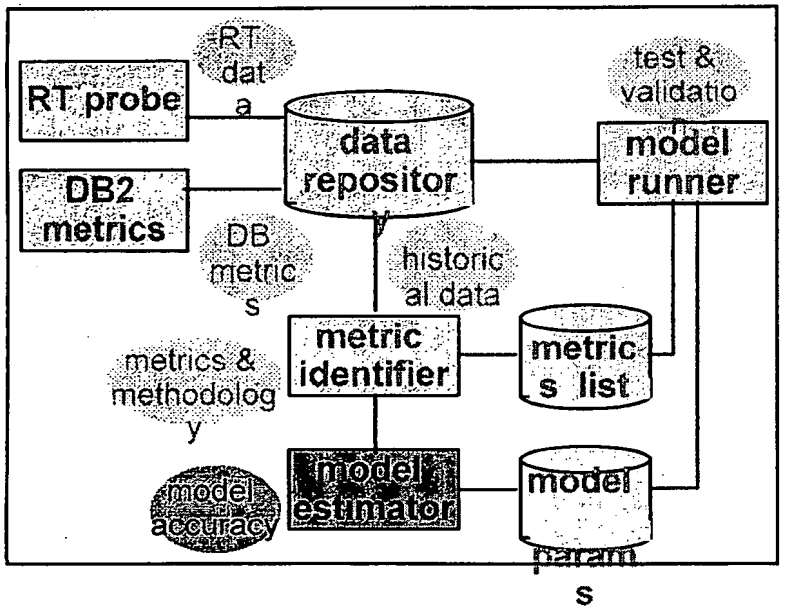
1. [25] Database Snapshot:Total sort time (ms)
2. [24] Database Snapshot:Total sorts
3. [10] Database Snapshot:Appls. executing in db manager currently
4. [58] Database Snapshot:Host execution elapsed time
5. [351] Tablespace Snapshot:Total buffer pool read time (ms):TS_ORDLN
6. [159] Bufferpool Snapshot:Total buffer pool read time (ms):BP_TEMP4K
7. [367] Tablespace Snapshot:Buffer pool data logical reads:IS_ORDLN

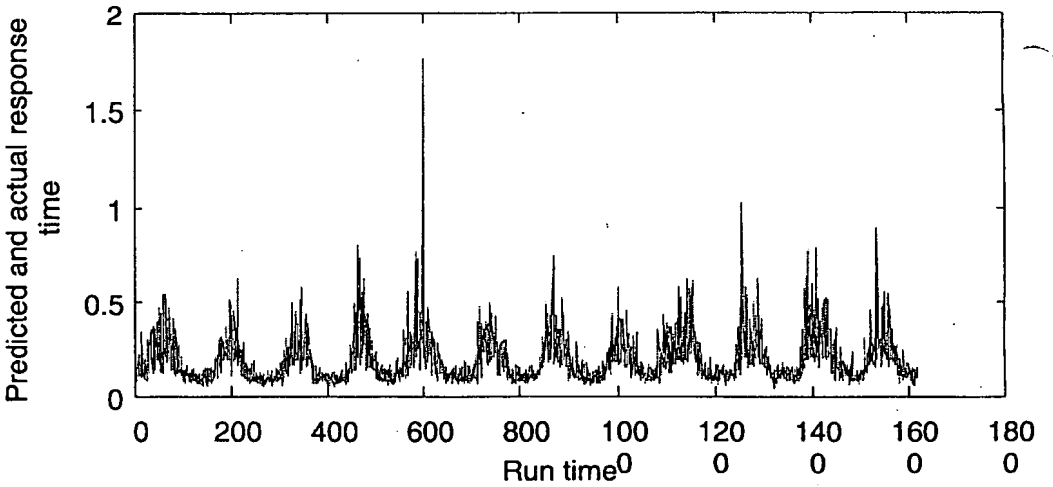


Metric discovery: build inference model for predicted response time- demo flow 5

- Build estimation model for response time with historical data
  - Use linear regression model
    - –dependency of model parameters on system and workload state not well characterized at this point
  - Other models possible - e.g. neural net

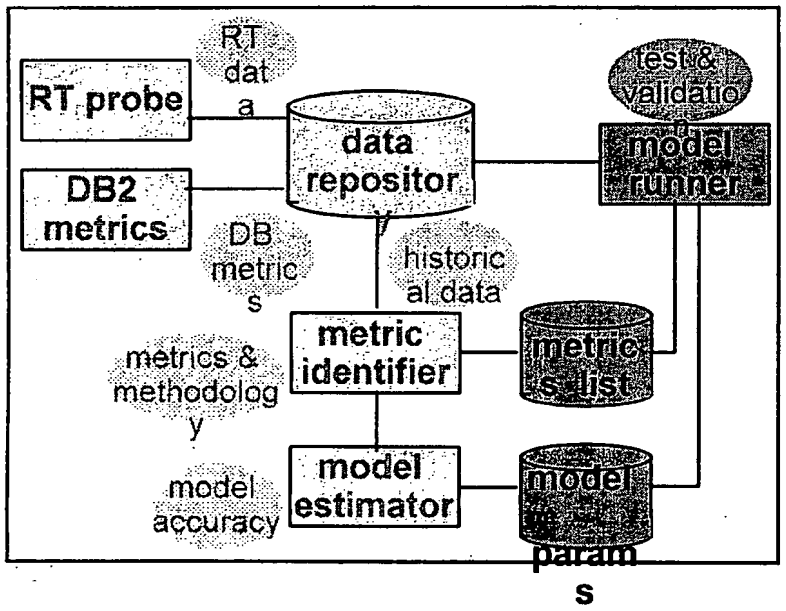


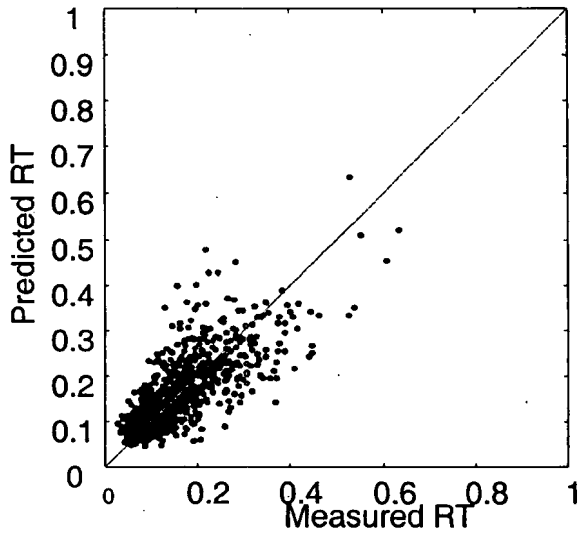


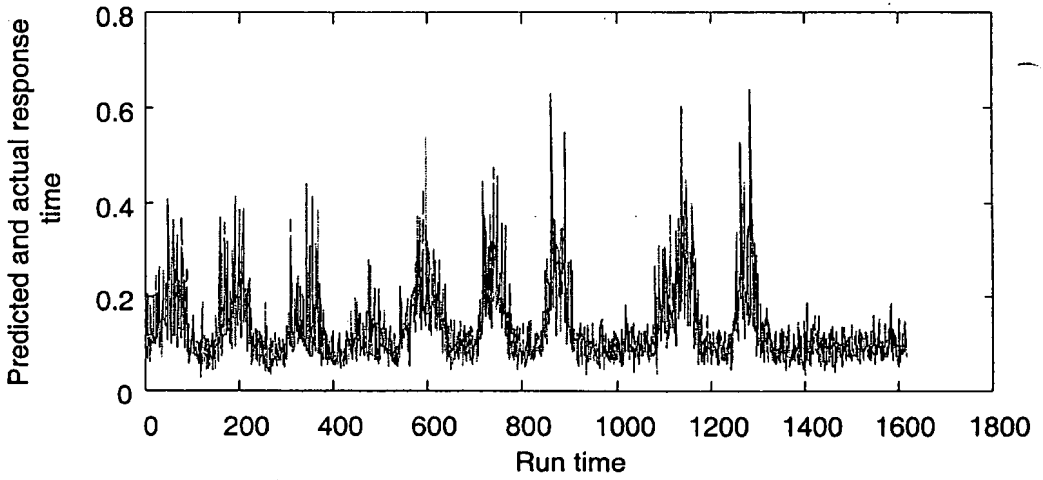


Metric discovery - demonstrate inference model time- demo flow 6

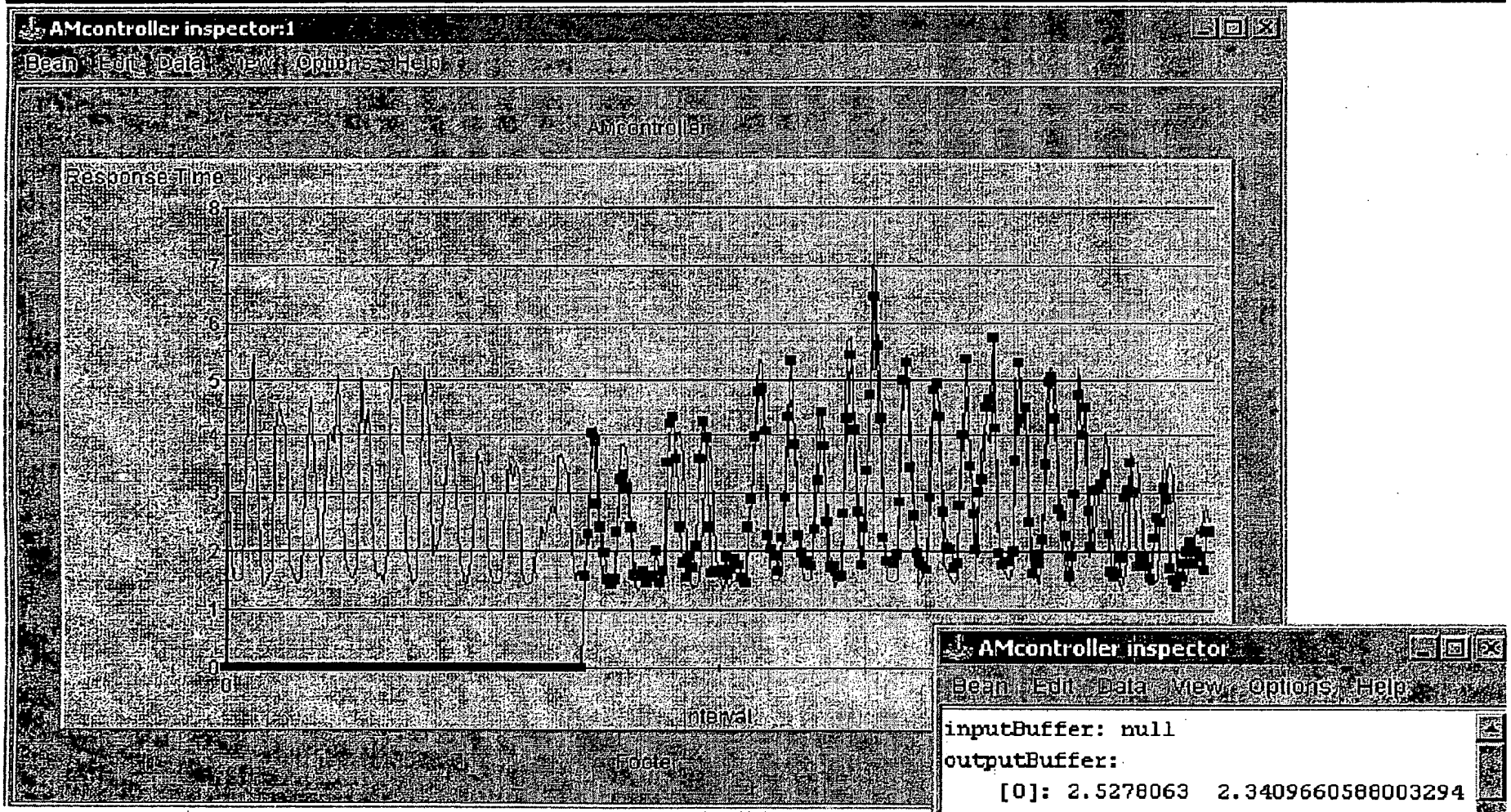
- Test inference model with saved test data
  - Not feasible to do in real time since RT probe data collected every 20 min only, and predicted RT reported every 2 min
  - robustness...
    - –across wider range of workload mix?
    - –system loads from other applications?







AM Operation



[10] Database Snapshot:Appls. executing in db manager currently
[299] Tablespace Snapshot:Total buffer pool read time (ms):TS_DATA
[39] Database Snapshot:Total buffer pool write time (ms)

RELATED PROCEEDINGS APPENDIX

None.